Corrections to "JPEG Dequantization Array for Regularized Decompression"

Wilfried Philips

***Abstract*—An error in the theoretical derivations leading up to the main equation of the above correspondence by Prost *et al.* is pointed out, and corrections are presented. Due to the error, the experimentally obtained "optimal" parameters are really only suboptimal.**

***Index Terms*—Artifact reduction, JPEG, regularization.**

I. Introduction

In the above mentioned correspondence[1] by Prost *et al.*, (7) relates the mean squared quantization error of a JPEG codec in the image domain to the corresponding error in the DCT domain. The original version of (7) should be replaced with the following correction:

$$\sum_{k=0}^{M-1}\sum_{l=0}^{M-1} N_i^2(k,\,l)Q^2(k,\,l) = \sum_{m=0}^{M-1}\sum_{n=0}^{M-1} n_i^2(m,\,n). \qquad (1)$$

The factor $Q^2(k,\,l)$ was missing in the original equation, but is required because the numbers $N_i(k,\,l)$ are *scaled* versions of the DCT coefficients of $n_i(m,\,n)$, which should be inversely scaled before applying Parseval's theorem.

The correction of (7) also impacts the derivations of (14), (16), (17), and (19). The most important of these equations is (19), which is the main result of the paper. The correct version of this equation is

$$\hat{Q}_i(k,\,l) = \frac{Q(k,\,l)}{1+\lambda_i D_{2M}^2(k,\,l)}. \qquad (2)$$

Note the absence of the factor $Q^2(k,\,l)$, which was present in the denominator of the original (19).

We note that $D_{2M}(k,\,l)$ is the Fourier-transform of a highpass filter $d(m,\,n)$ that in principle can be chosen freely (of course some filters will perform better than others). By choosing $D_{2M}(k,\,l)$ as $D_{2M}(k,\,l) = Q(k,\,l)D'_{2M}(k,\,l)$, where $D'_{2M}(k,\,l)$ is the value of $D_{2M}(k,\,l)$ in the original paper, and substituting in (2), we find that

$$\hat{Q}_i(k,\,l) = \frac{Q(k,\,l)}{1+\lambda_i Q^2(k,\,l)D'^2_{2M}(k,\,l)} \qquad (3)$$

which has the same functional form as the erroneous (19). This suggests that the numerical results in the original paper can be saved simply by choosing a different filter $D_{2M}(k,\,l)$. However, this is not correct because the parameter λ_i depends on $D_{2M}(k,\,l)$.

However, even though the original derivations result in a suboptimal solution, the experimental results indicate that the proposed method has some beneficial effects. This is somewhat surprising because the proposed method (which effectively scales the quantized DCT-coefficients before inversely transforming them) is basically an image-adapted lowpass filter implemented in the DCT-domain, which means that it is an *intrablock* filter, i.e., it does *not* filter over the block boundaries.

To conclude this comment, we present a possible explanation for the reduction of the block effects. First we note that scaling the quantized DCT coefficients is equivalent to filtering the even-symmetric extension of the reconstructed image block by a lowpass filter. In general, the even-symmetric extension of *any* image block is discontinuous at the block boundaries. In other words, such an extension contains artificial high-frequency components near the block boundaries. The proposed method smooths these artificial discontinuities and this effectively reduces the gradient magnitude near the block boundaries more than elsewhere. The corresponding extra smoothing near the block boundaries may account for the reduced block effects.

Manuscript received August 22, 1997; revised April 22, 1998. This work was supported by the Belgian Fund for Scientific Research (FWO-Vlaanderen) through a mandate of "Postdoctoral Research Fellow" and through the FWO Projects 39.0051.93 and 31.5831.95; and the Flemish Institute for the Advancement of Scientific-Technological Research in Industry (IWT) through Projects Tele-Visie (IWT 950202) and Samset (IWT 950204). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christine Podilchuk.

The author is with the Department of Telecommunications and Information Processing, University of Gent, B-9000 Gent, Belgium (e-mail: philips@telin.rug.ac.be).

Publisher Item Identifier S 1057-7149(98)08725-9.

[1] R. Prost, Y. Ding, and B. Atilla, *IEEE Trans. Image Processing*, vol. 6, pp. 883–888, June 1997.

Reproducing Kernel Hilbert Space Method for Optimal Interpolation of Potential Field Data

Jonathan Maltz, Robert De Mello Koch, and Andrew Willis

***Abstract*—The RKHS-based optimal image interpolation method, presented by Chen and de Figueiredo (1993), is applied to scattered potential field measurements. The RKHS which admits only interpolants consistent with Laplace's equation is defined and its kernel, derived. The algorithm is compared to bicubic spline interpolation, and is found to yield vastly superior results.**

***Index Terms*— Interpolation, potential field, reproducing kernel, Sobolev.**

I. Introduction

Many applications in geophysics, such as mineral exploration, rely on surveys of the Earth's local magnetic field for the detection of anomalies that reveal underlying geological features. In aeromagnetic imaging, field measurements are typically obtained from aeroplane-mounted magnetometers that sample the field at intervals along the flight path. Entire data sets, which consist of measurements collected over many parallel flight paths, typically exhibit large sample spacing irregularity and must undergo interpolation onto a regular grid to facilitate display and interpretation.

The techniques most commonly employed commercially rely on spline fitting algorithms which yield interpolants which are smooth,

Manuscript received October 5, 1996; revised March 5, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ping Wah Wong.

J. Maltz was with the Department of Electrical Engineering, University of the Witwatersrand. He is currently with the Department of Electrical Engineering and Computer Science, University of California, , Berkeley, CA 94720 USA (e-mail: jon@eecs.berkeley.edu).

R. De Mello Koch is with the Department of Physics, University of the Witwatersrand, Johannesburg 2050, South Africa.

A. Willis is with the Department of Electrical Engineering, University of the Witwatersrand, Johannesburg 2050, South Africa..

Publisher Item Identifier S 1057-7149(98)08719-3.

but which do not conform to a relevant physical model. The degree of smoothness must be manually set, and its specification is thus largely based on the subjective interpretation of the operator [1]. In addition, images output by these methods typically contain artifacts which render interpretation difficult. For example, long features perpendicular to the flight lines are typically split into several smaller features centered on each data point. This is termed the *herringbone effect.*

Spline methods other than standard bicubic spline fitting have been proposed. The minimum curvature method (MCM), produces interpolating surfaces that satisfy thin metallic sheet models [2]. This method exhibits similar limitations to bicubic spline techniques, and introduces the additional problem of surface oscillation in underconstrained areas. However, this problem, as well as the herringbone effect, has been ameliorated by the induction of variable spline tension [3], [1]. Unfortunately the user must still specify the spline tension factor.

The technique of kriging, or least-squares collocation, attempts to estimate the field at an interpolated point as a linear combination of that at neighboring sample points. Its advantage over a simple weighted-average method is the incorporation of a semivariogram describing spatial covariance. An anisotropic kriging algorithm has been proposed to overcome the herringbone effect in cases where the area surveyed contains anomalies which exhibit a preferred orientation [1]. The main drawback of this method is the need to first interpolate the data using an isotropic method in order to select an appropriate spatial model. This calls for a subjective interpretive step, after which the covariance matrix for the anisotropic kriging algorithm may be formulated.

In another approach to kriging, a three-dimensional (3-D) fractal stochastic model of the Earth's crust is developed for use in specifying the covariance matrix for the kriging algorithm. The resultant interpolants consequently have the same form of self-scaling power spectra as the crustal model, and are shown to possess a justifiable degree of smoothness [4].

None of the techniques described thus far exploit the important *a priori* knowledge that the interpolant of the potential field must lie on a harmonic surface, thus satisfying Laplace's equation [5]. Several that exploit this constraint are based on the equivalent source theorem which states that a two-dimensional (2-D) source distribution may be constructed at any depth to synthesize a field measured at the observation plane [5], [6]–[8]. Accordingly, interpolation is effected by creating an equivalent layer of regularly gridded sources whose contributions sum to the values at each measurement point. This method has been shown to yield results vastly superior to MCM [5], [9]. However, the results obtained depend strongly on the depth chosen for the equivalent layer. Too low a layer tends to precipitate instability, while layers which are too high induce aliasing [7]. The number of equivalent sources used must also be preselected. The severe computational requirements of the method have prompted the development of a technique which forms a smaller equivalent measurement set. The set is chosen such that the surface fitting these points fits the remaining points within a specified tolerance. This iterative method greatly reduces the computational burden [5].

The purpose of the work presented here is to develop a technique of potential field interpolation which produces an interpolant that satisfies Laplace's equation in an optimal sense and requires no heuristic selection of algorithm parameters. We show that this stipulation leads to an algorithm which produces interpolants of physically justifiable smoothness, and devoid of commonly encountered artifacts. Our approach is based on that of Chen and de Figueiredo (1993), who present a unified approach to optimal interpolation of images. The method they propose may be used to interpolate images generated by a processes which can be represented using a linear partial differential equation (LPDE) models.

We begin by illustrating how the interpolation problem at hand may be expressed in such form that the method of Chen and Figueiredo may be applied. Section III follows with a justification for the choice of the optimality criterion. The kernel of the relevant RKHS is then derived, and the numerical implementation of the algorithm discussed. The performance of the resulting algorithm is consequently evaluated.

II. Formulation of the Interpolation Problem

A. Stochastic Image Model

For reasons which will be later explained, it is useful to model the image as the output of a system having Gaussian white noise innovation. While [10], allow this innovation $u(t)$ to evolve in time, we consider a process $u(z)$ which evolves along the axis perpendicular to the interpolation plane.

Consider L, an LPDE operator which acts on a time varying image signal $f(x, y, z)$ to produce a zero-mean Gaussian white noise process. This may be expressed as [10]

$$Lf(x, y, z) = u(x, y, z). \tag{1}$$

Since L summarizes our knowledge of the image creation process, its form is known. Furthermore, the measurements yield values $\{a_{ij}\}_{j=1}^{n_i}{}_{,\,i=1}^{\,l}$, which correspond to points scattered, perhaps irregularly, in space. Thus [10], [11]

$$f(x_{ij}, y_{ij}, z_i) = a_{ij} \qquad j = 1, \cdots, n_i; \quad i = 1, \cdots, l. \tag{2}$$

It is required to reconstruct that $f(x, y, z)$ which satisfies both the image generation model of (1), and honors the measurement points according to (2).

B. Boundary Conditions

For generality, it is desirable that the most general knowledge of the image signal at its boundaries may be incorporated into the reconstruction. First, we stipulate that the image is defined in the standard Hilbert space, L_2, where Ω is the x-y domain containing the image, and I is the interval of support along the vertical (z) axis. The basis for the Hilbert space may be taken as any set of orthogonal complete functions which obey the boundary conditions

$$\begin{aligned} f(x, y, 0) = \psi_0^0(x, y), \quad f(x, y, Z) = \psi_0^Z(x, y), \\ f(x, y, z)|_\Gamma = \phi_0(x, y, z) \end{aligned} \tag{3}$$

where ψ^0, ψ^Z, and ϕ^0 are specified continuous functions; Γ is the spatial boundary of Ω, the x-y domain of the image; and Z is the position of interpolated plane along the z-axis, relative to the sources.

III. Minimum Energy Solution as Optimality Criterion

In (1), the operator L is known, while $f(x, y, z)$ can be specified at certain points and at the boundaries. The process $u(x, y, z)$ is, however, unknown, and so no unique solution exists.

In the solution of generic inverse problems, any one of several criteria may be employed to admit a unique solution. However, since out interpolant $\hat{f}$ exists within the RKHS F and is generated from a white noise innovation, by the representation theorem for RKHS's, minimization of $\|f\|_F^2$ is equivalent to that of $\|A^+ f\|_{L_2}^2$ in the general Hilbert space [12]. Identifying the Moore–Penrose pseudoinverse of the operator A, A^+, with L of (1), we see that the optimal interpolant f may be obtained *via a simple transformation between L_2 and the appropriate RKHS*. The choice of the minimum norm optimality criterion is thus by no means *ad hoc*, as it selects that solution *most* consistent with the model of (1).

IV. Optimal Image Interpolation Consistent with Laplace's Equation

It is required to find the self-reproducing kernel of that RKHS which produces an interpolant that honors the sample points and while also satisfying Laplace's equation. A further physical constraint is imposed which requires that the field tend to zero magnitude as the distance from the sources approaches infinity.

Rather than specify the initial, terminal and boundary conditions separately, we include them implicitly within the kernel function. This is simply due to the extreme simplicity of the BV's in the present case. Consequently, the interpolation goal becomes (cf. [10, (3.1)])

$$\min(L\hat{f}) \in \text{L2}(\Omega \times I) \iiint dx\ dy\ dz |L\hat{f}(x,y,z)|^2. \tag{4}$$

The initial-terminal conditions on z, of (3), when L is chosen as the 3-D Laplacian operator and the field is zero at infinity, are

$$Lf(x,y,0) = f(x,y,\infty) = Lf(x,y,\infty) = 0. \tag{5}$$

Comparing (3) and (5), it can be seen that no values at the boundary Γ are specified. As we will see, these values are not needed to find a unique interpolant $\hat{f}$. The final criterion which states that the interpolant must agree with the samples at each sample point is given simply by (2).

A. Derivation of Kernel

To solve the partial differential equation

$$L^\dagger L G(x,y,z) = \delta(x)\ \delta(y)\ \delta(z) \tag{6}$$

for the fundamental solution G, it is most convenient to work in Fourier space. Our conventions for the Fourier transform are

$$G(x,y,z) = \iiint dk_x\ dk_y\ dk_z\ e^{j(k_x x + k_y y + k_z z)} \cdot G(k_x,k_y,k_z) \tag{7}$$

where k_x, k_y, k_z are spatial wavenumbers.

We must now apply the operator $L^\dagger L$ to this function to find $G(k_x,k_y,k_z)$.

Since the operator L is self-adjoint, $L^\dagger L = L^2$, and application of the operator $\nabla^2(x,y,z)$ to (7) yields

$$L^2 G(x,y,z) = \iiint dk_x\ dk_y\ dk_z e^{j(k_x x + k_y y + k_z z)} \cdot G(k_x^2,k_y^2,k_z^2)[k_x^2 + k_y^2 + k_z^2]^2. \tag{8}$$

Comparing this expression with the Fourier transform of the Dirac delta distribution

$$\delta(x)\ \delta(y)\ \delta(z) = \iiint \frac{dk_x}{2\pi}\ \frac{dk_y}{2\pi}\ \frac{dk_z}{2\pi}\ e^{j(k_x x + k_y y + k_z z)} \tag{9}$$

it is obvious that

$$G(x,y,z) = \iiint \frac{dk_x}{2\pi}\ \frac{dk_y}{2\pi}\ \frac{dk_z}{2\pi}\ \frac{e^{j(k_x x + k_y y + k_z z)}}{(k_x^2 + k_y^2 + k_z^2)^2}. \tag{10}$$

Equation (10) gives the fundamental solution (Green's function) of the differential equation (6). Since we have already addressed the BV and ITV conditions, the kernel is equal to the fundamental solution and we define $K(x,y,z) = G(x,y,z)$ (cf. [10, (3.10)]).

Only harmonic functions may be represented in terms of this basis, in the second Sobolev RKHS which it spans. In addition, all functions represented in terms of this basis vanish as the vertical (z) distance from the sources approaches infinity.

B. Integrating the Kernel Analytically

As it stands, the function G given in (10) is not defined. There is a singularity lying on the integration contour at $k_x = k_y = k_z$. To define this, we need to stipulate how to treat this pole. We define the integral by picking up the residue of the pole at $k_z = j\sqrt{k_x^2 + k_y^2}$.

We require k_x and k_y to be real. This prescription is indicated by the boundary conditions.

After defining the variables $k = \sqrt{k_x^2 + k_y^2}$ and $\theta = \arctan(k_y/k_x)$, we perform the factorization $k_x^2 + k_y^2 + k_z^2 = [k_z - j\omega_k][k_z + j\omega_k]$, where $\omega_k = k_x^2 + k_y^2$. We then integrate over k and then θ yielding the final expression for the reproducing kernel as

$$G = \frac{1}{4(2\pi)^2}\left[\int_0^{2\pi} d\theta[\ln(-(j\cos(\theta)x + j\sin(\theta)y - z)) \times [(-(j\cos(\theta)x + j\sin(\theta)y - z)) + z] - (-(j\cos(\theta)x + j\sin(\theta)y - z))]\right]. \tag{11}$$

V. Algorithm Implementation

[10, Th. A], states that the optimal solution to the interpolation problem may be expressed as

$$v^* = \sum_{i=i}^{l} \sum_{j=1}^{N_i} b_{ij} K(x_{ij}, y_{ij}, z_i, x, y, z) \tag{12}$$

where the coefficients $b_{ij}\ {}^{n_i}_{j=1},{}^{L}_{i=1}$ are uniquely determined by interpolation conditions of consistency with the N_i measurement points and specified boundary values.

In our case, where the BV's and ITV's are included in the kernel, the b_{ij} may be calculated as the solution to the equation

$$\sum_{i=1}^{L} \sum_{j=1}^{N_m} b_{ij} K(x_{pq} - x_{ij}, y_{pq} - y_{ij}, z_p - z_i) = a_{pq} \quad q = 1,\cdots,n_p; \quad p = 1,\cdots,l \tag{13}$$

where N_m denotes the number of measurement points in the image at the measurement plane $z = L$. Equation (13) may also be expressed in matrix form as

$$K_{(pq,ij)} b_{ij} = a_{pq}. \tag{14}$$

In summary, the following matrix expression yields the values $\hat{v}_{rs}$ at the interpolated points:

$$\hat{v}_{rs} = K_{(rs,ij)} K^{-1}_{(pq,ij)} a_{pq}. \tag{15}$$

From this equation, it is clear that the sample space is first mapped into the range of $K^{-1}_{(pq,ij)}$, a matrix of kernel functions whose arguments involve the differences between the measurement point coordinates, and then into the range of $K_{(rs,ij)}$, whose arguments consist of the differences between the N_i interpolated point coordinates and those of the measurement points.

The algorithm requires the calculation of the elements of the $N_m \times N_m$ matrix $K^{-1}_{(pq,ij)}$ and the $N_i \times N_m$ matrix $K_{(rs,ij)}$. Each element requires the numerical integration of the kernel function, and the filling of these matrices is consequently very computationally intensive. An adaptive recursive Newton–Cotes eight panel rule is used to perform the integration.

The matrix inversion required in (15) is expensive in terms of both computation and storage if the number of measurement

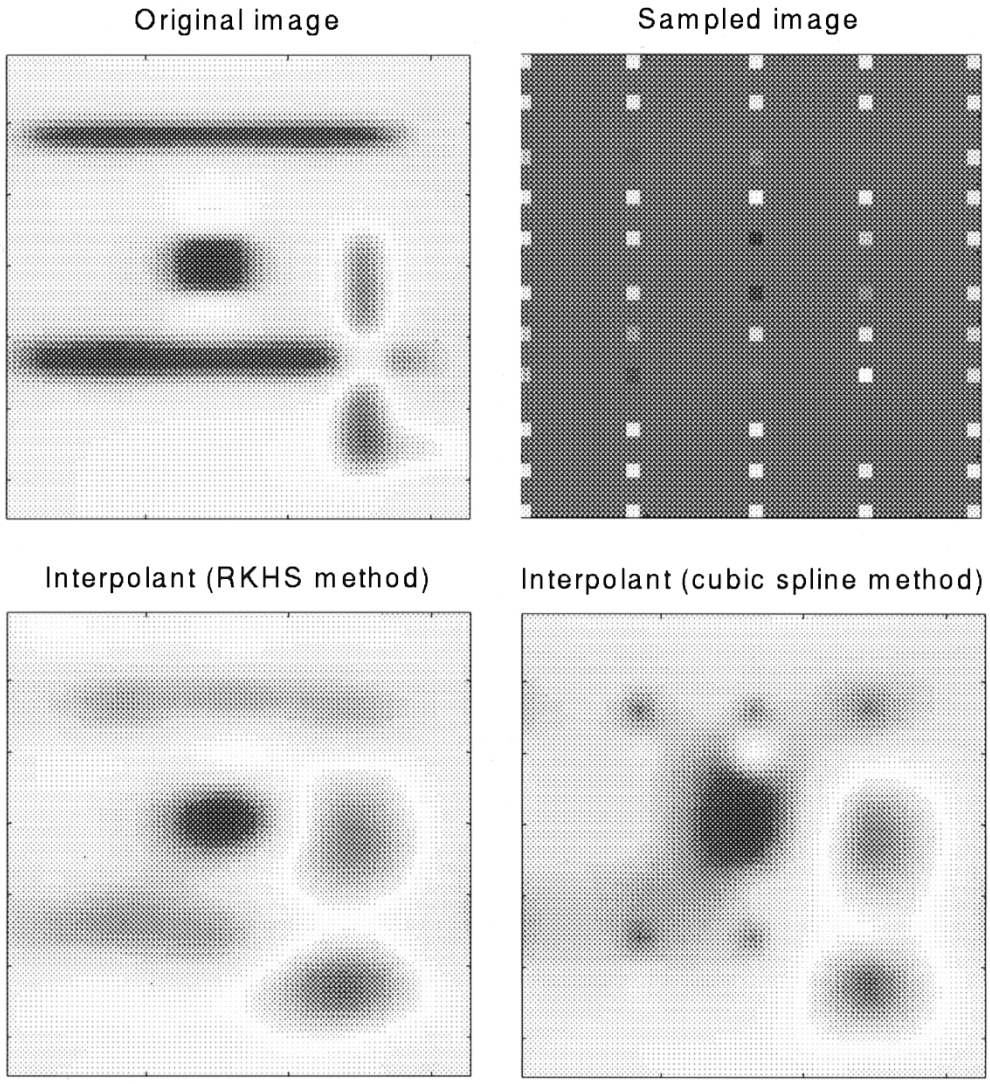

Fig. 1. Test I: With regular spacing, RKHS method yields superior results.

TABLE I
PARAMETERS FOR TEST I

Parameter	*Value*
Data set size (pixels)	65 x 65
Total number of samples used	55
Total number of points in interpolant	4225
Interpolated image size (pixels)	65 x 65
Results depicted in figure	1

TABLE II
PARAMETERS FOR TEST II

Parameter	*Value*
Data set size (pixels)	65 x 65
x co-ordinate of sample standard deviation	5%
y co-ordinate of sample standard deviation	7.5%
Total number of samples used	55
Total number of points in interpolant	4225
Interpolated image size (pixels)	65 x 65
Results depicted in figure	2

points used is very large. Overall, computational burden increases as $O(N_m^3 + N_m(1 + N_i))$.

VI. EXPERIMENTAL EVALUATION OF THE INTERPOLATION METHOD

Synthetically generated magnetic potential field data are used to evaluate the performance of the RKHS interpolation method. The images consist of the combined contributions of the fields of vertical sided anomalies of infinite depth extent. The formulae applied for the generation of this data are due to [13]. The data consist of points irregularly sampled from a 65 $\times$ 65 pixel image.

The performance of a commercial geophysical interpolation package employing bicubic spline methods is used as a reference for this comparative study.

Figs. 1 (Test I) and 2 (Test II), compare RKHS and cubic spline interpolants for test parameters specified in Tables I and II. Visual inspection reveals that the RKHS interpolation algorithm produces superior results. This is borne out by the sum of squared differences statistics in Table III.

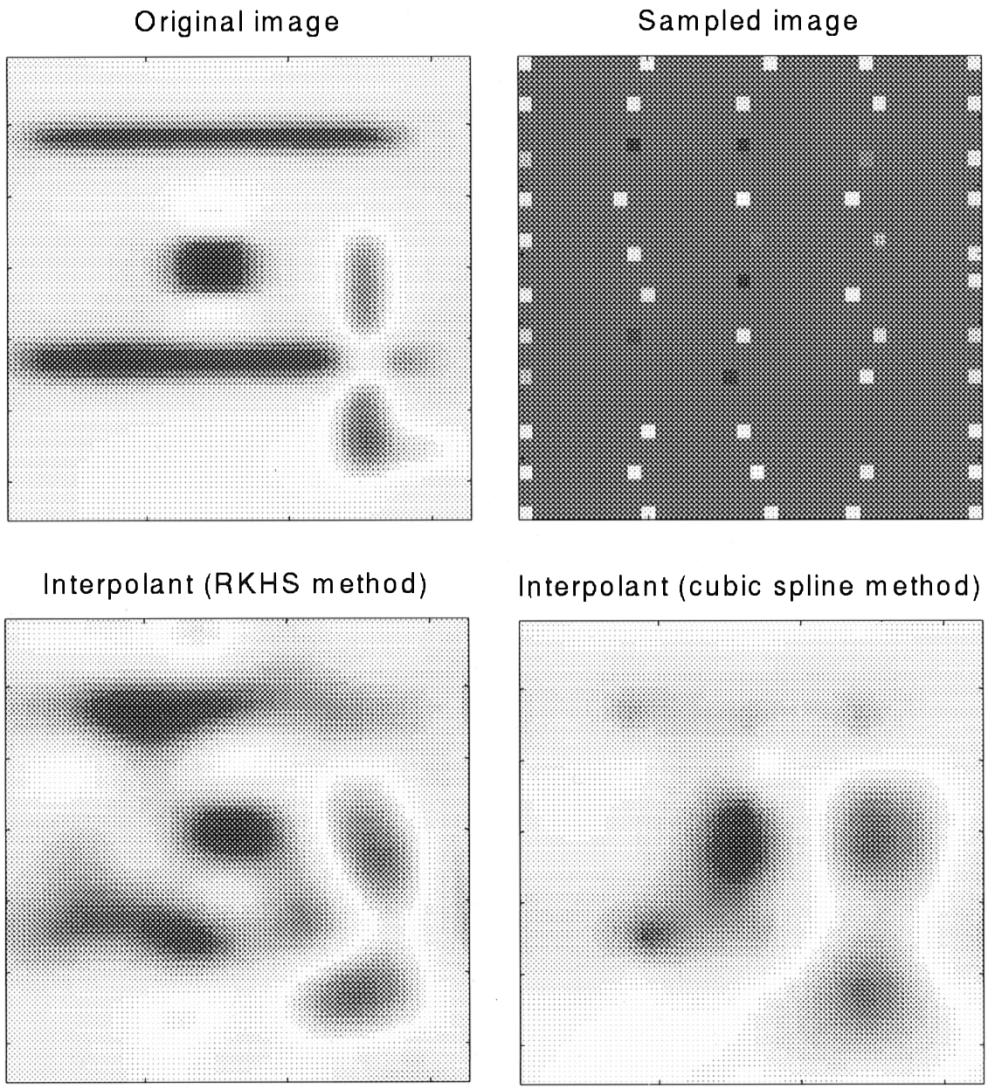

Fig. 2. Test II: With irregular spacing, RKHS method yields vastly superior results.

TABLE III
RESULTS FOR COMPARATIVE ANALYSIS (TEST II)

Figure	*SSD (spline)*	*SSD (RKHS)*
Figure 1	2.12E-2	1.26E-2
Figure 2	2.43E-2	1.51E-2

In this example, the RKHS method requires 1270 Mflops, executing in a time of 102 min on a Sun Ultra II under Matlab 5.

VII. CONCLUSION

The optimal interpolation method originally introduced by Chen and de Figueiredo has been applied, using the nonseparable 3-D Laplace operator, to the problem of potential field interpolation.

The utilization of the *a priori* knowledge that the measurements represent samples of an harmonic field, leads to interpolant qualitatively superior to bicubic spline methods, where only the smoothness of the interpolant is guaranteed. The resulting algorithm requires no manual specification of either smoothness, a dominant orientation model, or equivalent source layer depth.

It is thought that the major disadvantage of this method, its computational intensiveness, might be addressed through reduction to a smaller equivalent data set as advocated by Mendonca and Silva. This technique is likely to prove helpful in the present case, as it has ameliorated the very similar computational problems encountered in the implementation of the equivalent source method [5].

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Reid of Getech, University of Leeds, U.K., for his assistance in the development of this technique.

REFERENCES

[1] R. O. Hansen, "Intepretative gridding by anisotropic kriging," *Geophysics*, vol. 58, pp. 1491–1497, 1993.
[2] I. C. Briggs, "Machine contouring using minimum curvature," *Geophysics*, vol. 39, pp. 39–48, 1974.
[3] W. H. F. Smith and P. Wessel, "Gridding with continuous splines in tension," *Geophysics*, vol. 55, pp. 293–305, 1990.
[4] M. Pilkington, M. E. Gregotski, and J. P. Todoeschuck, "Using fractal crustal magnetization models in magnetic interpretation," *Geophys. Prospect.*, vol. 42, pp. 677–692, 1994.

[5] C. A. Mendonca and J. B. C. Silva, "The equivalent data concept applied to the interpolation of potential field data," *Geophysics*, vol. 59, pp. 722–732, 1994.
[6] F. S. Grant and G. F. West, *Interpretation Theory in Applied Geophysics*. New York: McGraw-Hill, 1965.
[7] C. N. G. Dampney, "The equivalent source technique," *Geophysics*, vol. 34, pp. 39–53, 1969.
[8] L. Cordell, "A scattered equivalent source method for interpolation and gridding of potential field data in three dimensions," *Geophysics*, vol. 57, pp. 629–636, 1992.
[9] C. A. Mendonca and J. B. C. Silva, "Interpolation of potential field data by equivalent layer and minimum curvature: A comparative analysis," *Geophysics*, vol. 60, pp. 399–407, 1995.
[10] ____, "Interpolation of potential field data by equivalent layer and minimum curvature: A comparative analysis," *Geophysics*, vol. 60, pp. 399-407, 1995.
[11] G. Chen and R. J. P. de Figueiredo, "A unified approach to optimal image interpolation problems based on linear partial differential equation models," *IEEE Trans. Image Processing*, vol. 2, pp. 41–49, 1993.
[12] R. J. P. de Figueiredo and G. Chen, "PDL_g splines defined by partial differential operators with initial and boundary value conditions," *SIAM J. Numer. Anal.*, vol. 27, pp. 519–528, 1990.
[13] T. Kailath, "RKHS approach to detection and estimation problems—Part I: Deterministic signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 530–549, 1971.
[14] B. Bhattacharyya, "Magnetic anomalies due to prism-shaped bodies with arbitrary polarization," *Geophysics*, vol. 29, pp. 517–531, 1964.

MAP Image Restoration and Segmentation by Constrained Optimization

Stan Z. Li

***Abstract*—The combinatorial optimization problem of MAP estimation is converted to one of constrained real optimization and then solved by using the proposed augmented Lagrange–Hopfield (ALH) method. The ALH effectively overcomes instabilities that are inherent in the penalty method or the Lagrange multiplier method in constrained optimization. It produces good solutions with reasonable costs.**

***Index Terms*—Augmented Lagrange method, combinatorial optimization, constrained optimization, graded Hopfield networks, Markov random fields (MRF's), maximum *a posteriori*.**

I. Introduction

The aim of image restoration is to recover a degraded image and that of image segmentation is to partition an image into regions of similar image properties. Efficient restoration and segmentation are very important for numerous image analysis applications. Both problems can be posed generally as one of image estimation where the underlying image or segmentation map is to be estimated from the degraded image. Due to various uncertainties, an optimal solution is sought. A popular optimality criterion is the maximum *a posteriori* (MAP) probability principle in which both the prior distribution of the true image class and the conditional (likelihood) distribution of the data are taken into account. Contextual constraints, i.e., constraints between pixels, are important in image analysis. Markov random fields (MRF's) or equivalently Gibbs distributions provide a convenient tool for modeling prior distributions of images which encode contextual constraints. Maximizing the posterior is equivalent to minimizing the energy function in the Gibbs distribution. The MAP principle and MRF together form the MAP-MRF framework [1], [2].

When the pixels of the image to be recovered take discrete values, as is the case dealt with in this paper, the minimization is combinatorial. Discrete optimization methods often used in statistical image analysis include the iterative conditional modes (ICM) [3] and simulated annealing (SA) [1], [4]. Other discrete algorithms include the highest confidence first (HCF) [5]. The deterministic algorithms of ICM and HCF quickly converge to a local energy minimum but are dependent largely on the initial configuration. The stochastic SA with a slow enough schedule finds a global solution with probability approaching one but is well-known to be expensive.

A combinatorial optimization can often be converted into a constrained real optimization with equality and inequality constraints. The penalty and the Lagrange multiplier methods can be used for coping with equality constraints and the barrier method for coping with inequality constraints. However, the penalty method suffers from the ill-conditioning and the Lagrange method suffers from the zigzagging problem [6]. The augmented Lagrange (AL) method [7] combines both the Lagrange and the penalty methods and effectively overcomes the associated problems. In AL, the relative weight for the penalty terms need not be infinitely large, this not only overcoming the ill-conditioning problem but also beneficial for obtaining better quality solutions because the relative importance of the original objective function is more emphasized; on the other hand, its use of quadratic penalty terms "convexifies" and hence stabilizes the system, overcoming the zigzagging problem [6].

Mean field annealing (MFA) [8] provides still another continuous method. Assuming that the minima of the original energy and the corresponding mean field effective energy coincide, the MFA aims to approximate the global minimum of the original energy by tracking that of the effective energy with decreasing temperature. A recent analysis shows that the effective energy of MFA is identical to a combination of the original energy, a particular barrier term and a standard Lagrange term [9].

In this work, we present another deterministic method, called the augmented Lagrange–Hopfield (ALH) method, for the combinatorial optimization in the MAP-MRF image restoration and segmentation. In solving the converted constrained real optimization, the ALH method uses the augmented Lagrangian multiplier method [7] to satisfy the equality constraints and the Hopfield network encoding [10] to impose the inequality constraints. The use of AL effectively overcomes instabilities inherent in the penalty method and the Lagrange multiplier method. The resulting algorithm solves a system of differential equations. Experimental results in both image restoration and segmentation are shown to compare the ALH method with the ICM, HCF, and SA. The results show that the ALH outperforms ICM and HCF, and is comparable to SA in terms of the solution quality; it quickly yields a good solution after a dozen of iterations, a number similar to that required by ICM and HCF but much smaller than SA. A discussion on MFA results is also provided.

The rest of the correspondence is organized as follows. Section II describes the ALH method after introducing the MAP-MRF formula-

Manuscript received October 16, 1995; revised April 7, 1998. This work was supported by NTU AcRP Projects RG43/95 and RG51/97. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Reginald L. Lagendijk.

The author is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: szli@szli.eee.ntu.edu.sg).

Publisher Item Identifier S 1057-7149(98)08722-3.